DOCUMENT RESUME

ED 138 635 TM 006 262

AUTHOR Scheetz, James P.; Forsyth, Robert A.
TITLE Assigning Items to Subtests in Multiple Matrix Sampling.
PUB DATE [Apr 77]
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Error Patterns; *Item Sampling
IDENTIFIERS *Multiple Matrix Sampling

ABSTRACT

The choice of design parameters (e.g., number of subtests, number of items per subtest, and number of examinees per subtest) can be controlled by the test constructor in a multiple matrix sampling evaluation. The purpose of this study was to determine empirically which combination of the above parameters produces the smallest standard errors of the estimated mean and variance when certain assumptions are violated. The results indicate that multiple matrix sampling is a robust procedure even though proposed guidelines which should be followed to minimize the standard error terms are ignored. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED138635

29.10

ASSIGNING ITEMS TO SUBTESTS IN MULTIPLE MATRIX SAMPLING

James P. Scheetz
Health Sciences Center
University of Louisville
Louisville, Kentucky 40201

and

Robert A. Forsyth
University of Iowa
Iowa City, Iowa 52242

Presented at the annual meeting of the
American Educational Research Association
April 4-8, 1977
New York, N.Y.

TM006 262

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The use of multiple matrix sampling (MMS) techniques for program evaluation purposes has been accepted by educational evaluators as a means of reducing the time and cost required to perform program evaluations (Knapp, 1972). The usual MMS procedure involves the administration of a sample of items drawn from a larger item universe to a sample of respondents drawn from a population of respondents. On the basis of this information estimates of population parameters (usually the mean and variance) are obtained and used as part of the evaluation data. It is desirable, of course, to estimate these parameters as accurately as possible.

Many investigators have attempted to identify procedures for selecting the MMS design that will provide the most accurate estimates of the population parameters [See, for example, Shoemaker (1970a, 1970b, 1971, 1972, 1973), Knapp (1972), and Barcikowski (1972, 1974)]. Shoemaker (1973) has performed numerous investigations of this problem and has formulated a set of guidelines for designing MMS studies that presumably will minimize the standard errors of the MMS estimates. These guidelines are intended to help the evaluator select t (the number of subtests), k (the number of items per subtest), and n (the number of respondents per subtest). Most of these guidelines evolved from a large empirical base; however, some are based on relatively limited evidence. In this study additional data related to two of Shoemaker's guidelines were gathered. The effects of rerandomizing the item universe after all items have appeared on one subtest as compared to not rerandomizing the items were also examined. Specifically, the following

three questions were investigated in this study:

1. When the product tk>K, does rerandomization or nonrerandomization of the item universe after single exhaustion has occurred result in more accurate estimates of the mean and variance?

This question is not directly related to any of Shoemaker's guidelines, but is indirectly related to the guideline which states that items and examinees should be sampled without replacement from the universe and population. Rerandomization of the item universe after single exhaustion involves randomizing the item universe before further sampling of items occurs. Nonrerandomization of items involves the repetition of the items among subtests in the same order as they appeared during single exhaustion of the item universe.

The rerandomization question has not been previously investigated. It is of interest because the investigator has the option of rerandomizing the items or administering them in the same order in which they were administered before single exhaustion occurred.

2. If tk>K, are more accurate estimates of the mean and variance obtained when tk is an integer multiple of K or when tk is not an integer multiple of K?

This question is related to Shoemaker's guideline which states that when tk exceeds K, (a) tk should be an integer multiple of K, and (b) all items should appear tk/K (i.e., an equal number of times) among the t subtests.

In an early study, Shoemaker (1970b) sampled items both with and without replacement among subtests. For those sampling plans in which tk exceeded K, the items were replaced in the item pool after each sample was drawn. The conclusion resulting from this study was that the total number of observations, O, (O = tkn), rather than the particular

values of t, k, and n is the important variable in reducing the standard errors of the estimated mean [$SE(\hat{\mu})$] and variance [$SE(\hat{\sigma}^2)$]. In a subsequent study, Shoemaker (1971) constructed sampling plans such that tk exceeded K and was not an integer multiple of K. Shoemaker modified his previous conclusion by indicating that the selection of values for t and k as well as the total number of observations should be considered when estimating the mean. Shoemaker (1972) subsequently stated that had the assignment of items to subtests been more systematic in these previous studies, the results probably would have agreed more closely with the conclusion of this later investigation which showed that when tk>K, the standard errors of the estimated mean and variance were smaller when tk was an integer multiple of K than when tk was not an integer multiple of K.

3. For a given normative distribution with 0 fixed, what type of design produces the most accurate estimates of the mean and variance?

This question is related to the guideline which states that making t large and k small when sampling from skewed distributions and vice versa when sampling from a normal distribution is the most effective means of reducing the estimated standard error terms.

Shoemaker's early investigations (1970a, 1970b) led to the conclusion that the parameters of the sampling plan, with the exception of 0, were not important variables to manipulate as a means of reducing $SE(\hat{\mu})$ and $SE(\hat{\sigma}^2)$. Further investigations by Shoemaker (1971, 1972, 1973) resulted in the recommendation that increasing t is the most effective means of reducing the $SE(\hat{\mu})$ and $SE(\hat{\sigma}^2)$ when sampling from negatively or positively skewed distributions whereas increasing k is most effective when sampling from a normal distribution.

Barcikowski (1972, 1974) and Moy and Barcikowski (1974) have

investigated this problem and found that it may not be possible to obtain the most accurate estimates of μ and σ^2 with the same sampling plan. Their findings suggest that the most effective method for estimating μ is to use a sampling plan with a large value of t and make k as small as possible. When estimating σ^2 they content that item discrimination indices (i.e., the biserial correlation between item and total test score) should be considered. For tests with biserial correlations in the range .05 - .55, the value of k should approach K/2. With tests of higher item discrimination, the number of items per subtest may be reduced.

Methodology

The method of investigation used in this study was the post-mortem approach in which the parameters of the data base are computed, samples are then drawn from this data base and the sample estimates are compared with the data base values. In this investigation only the mean and variance were estimated, although higher order moments could have been estimated.

Description of the Data Base

The items used in this study were taken from two subtests of the Iowa Tests of Educational Development (ITED): mathematics (36 items) and vocabulary (40 items). These two subtests were chosen because the distributions of scores on these subtests are known to be different. In addition, the items on each subtest are independent since the response to a particular item is not dependent on the response to other items.

During 1971, 16,819 ninth grade Iowa students took the math test (Form X-6). A 1 in T systematic sample was drawn to give a data base with N = 600. For this data base of 600 the following values were found:

$\mu = 11.623$, $\sigma^2 = 30.385$, $\sigma = 5.512$, skewness = .865, and kurtosis = 3.680. The range was 32 with a minimum score of one and a maximum score of 33. The reliability (KR 20) for the data base was .779. The distribution was positively skewed and leptokurtic.

The second data base was derived from the scores of 13,821 eleventh grade students in Iowa who took the vocabulary test (Form X-6) in 1973. A 1 in T systematic sample was drawn to construct a data base with N = 600. For this data base the following values were found: $\mu = 22.682$, $\sigma^2 = 88.328$, $\sigma = 9.398$, skewness = -.076, and kurtosis = 1.971. The range was 37 with a minimum score of three and a maximum score of 40. The reliability (KR 20) for the data base was .923. The distribution was slightly negatively skewed and platykurtic.

Description of Sampling Plans

Simple random sampling was the only sampling procedure used in this study. The sampling plans for the math test are listed in Table 1 and those for the vocabulary test are listed in Table 2. The column labeled RANDOM indicates whether or not the item universe was rerandomized after single exhaustion of the item universe occurred. The column labeled NREPS specifies the number of times each sampling plan was replicated.

Table 1 - Here

Table 2 - Here

Methods of Comparison

Each replication of a sampling plan resulted in a pooled estimate of the mean and variance. The pooled estimate of any parameter is given

by Shoemaker (1973, p. 38) as

$$\hat{\zeta}_p = \frac{\sum_{s=1}^{t} O_s \hat{\zeta}_s}{\sum_{s=1}^{t} O_s} \qquad (1)$$

where $\hat{\zeta}_s$ is the estimate of parameter ζ for subtest s, t is the number of subtests and $O_s = n_s k_s$, the number of observations obtained from subtest s. After the estimates of μ and σ^2 were obtained for each replication, estimated standard errors of these quantities were calculated using the following formula:

$$SE(\hat{\zeta}_p) = \left[\frac{\sum_{p=1}^{NREPS} (\hat{\zeta}_p - \bar{\hat{\zeta}}_p)^2}{NREPS} \right]^{\frac{1}{2}} \qquad (2)$$

where $\bar{\hat{\zeta}}_p$ is the mean over replications of the $\hat{\zeta}_p$ values.

Two other statistics, Ψ_1 and Ψ_2, involving the average absolute deviations of $\hat{\mu}_p$ and $\hat{\sigma}^2_p$ from μ and σ^2 of the data base were calculated in order to compare the sampling plans. For the mean,

$$\Psi_1 = \frac{\sum_{p=1}^{NREPS} |\mu - \hat{\mu}_p|}{NREPS} \qquad (3)$$

was computed where μ is the mean of the data base and $\hat{\mu}_p$ is the mean obtained from replication p of a particular sampling plan. This statistic is of interest because it indicates how accurately the mean computed by the MMS procedure estimates the true mean of the data base. A similar statistic, Ψ_2, was computed to measure the degree of accuracy of the

estimated variance. The statistic is

$$\Psi_2 = \frac{\sum_{p=1}^{NREPS} |\sigma^2 - \hat{\sigma}_p^2|}{NREPS} \qquad (1)$$

where σ^2 is the variance of the data base and $\hat{\sigma}_p^2$ is variance obtained from replication p of a given sampling plan.

The standard error terms show how closely the estimates of μ and σ^2 resulting from each replication cluster around the average of these estimates. The average absolute deviations indicate how accurately $\hat{\mu}_p$ and $\hat{\sigma}_p^2$ estimate the data base values of μ and σ^2, respectively.

Results and Discussion

The results for the math test are presented in Table 3 and those for the vocabulary test are contained in Table 4. In these tables the column labeled RAND indicates whether or not the item universe was re-randomized after single exhaustion occurred.

Table 3 - Here

Table 4 - Here

1. When the product tk>K, does rerandomization or non-rerandomization of the item universe after single exhaustion has occurred result in more accurate estimates of the mean and variance?

Data for sampling plans 6 through 17 of Table 3 and plans 5 through 18 of Table 4 were analyzed to help answer this question. There were six possible comparisons for the math test (e.g., plans 6 and 7) and

seven possible comparisons for the vocabulary test (e.g., plans 5 and 6).

Table 5 shows the number of times each method (rerandomization and nonrerandomization) resulted in the lower criterion value. For example, when considering the math test and $SE(\hat{\mu}_p)$ as the criterion, nonrerandomization of the items after single exhaustion resulted in a smaller value of $SE(\hat{\mu}_p)$ in four of the six comparisons. The data in Table 5 suggest that there is a tendency for the plans not involving rerandomization to yield more accurate estimates of μ and σ^2.

Table 5 - Here

While any conclusions based on these limited results must be considered cautiously, the implications of such results seem clear. When administering an instrument using MMS procedures, the item universe does not need to be rerandomized after single exhaustion has occurred. Estimates of μ and σ^2 may be obtained by administering the items in the same order as they were originally administered. This results in increasing the number of respondents per subtest since the order of the items is not changed for succeeding subtests.

2. If $tk>K$, are more accurate estimates of the mean and variance obtained when tk is an integer multiple of K or when tk is not an integer multiple of K?

On the basis of the previous results which indicated that rerandomization of the item universe tended to yield larger values of $SE(\hat{\mu})$ and $SE(\hat{\sigma}^2)$ than nonrandomization, only the results from the nonrerandomized plans will be considered and they will be divided into two groups: those with $t< k$ and those with $t>k$. These results are presented in Table 6.

Table 6 - Here

For the plans with $t < k$, the results for both the math and vocabulary tests indicate no clear trends when $SE(\hat{\mu}_p)$ and Ψ_1 are the criteria. Likewise, when $SE(\hat{\sigma}^2_p)$ and Ψ_2 are used as criteria, no trends are apparent for either the math or vocabulary tests. For some comparisons, the criterion value is less for the plans with $tk = aK$ where a is an integer. However, in other comparisons, the plans with $tk = aK$ have the greater criterion value. While the results of the study are somewhat limited, they do indicate the need for more research on this issue.

3. For a given normative distribution with O fixed, what type of design produces the most accurate estimates of the mean and variance?

When $tk = K$, it is known that the most accurate estimate of μ is obtained when $t = K$ and $k = 1$. (See Lord and Novick, 1968, p. 256). In addition, there is considerable empirical evidence which indicates that more accurate estimates of the mean are obtained by making t large relative to k. [See, for example, Barcikowski (1972, 1974), Moy and Barcikowski (1974), and Shoemaker (1971)]. The limited results of the current study offer only partial support for this conclusion. However, as noted above, the previous evidence seems fairly conclusive.

When $tk \neq aK$, the results of this study provide no support for making t large relative to k when estimating μ. The results for the math test indicate that making t large relative to k may reduce $SE(\hat{\mu}_p)$, but the results from the vocabulary test do not support this finding.

When $tk = 2K$, the results for the math test clearly indicate that t should be large relative to k on the basis of both $SE(\hat{\mu}_p)$ and Ψ_1. This result is similar to that found for the vocabulary test when

tk = 2K in that the $SE(\hat{\mu}_p)$ values are slightly smaller when t is large relative to k. The evidence from this study suggests that when tk = 2K, making t large relative to k will result in smaller values of $SE(\hat{\mu}_p)$ and Ψ_1.

The above results provide limited support for Shoemaker's (1973, p. 22) guideline that increasing t is the most effective means of reducing $SE(\hat{\mu}_p)$ when sampling from a skewed distribution. This result is also in agreement with the findings of Moy and Barcikowski (1974) who recommended that t should be large and k should be small when estimating μ.

When choosing among sampling plans with different values of t and k for estimating the variance, an analysis of $SE(\hat{\sigma}^2_p)$ for both the math and vocabulary tests and the three categories, tk = K, tk ≠ aK, and tk = 2K, provides no indication as to whether t should be large and k should be small or vice versa. No trends of any kind emerged which might provide an insight into the question regarding the choice of t and k. Likewise, when Ψ_2 is the criterion, the results from both tests do not indicate the optimal values of t and k for any of the three categories tk = K, tk ≠ aK, and tk = 2K.

The above results provide little support for Shoemaker's conclusion that when sampling from a normal distribution, increasing k is the most effective means of reducing $SE(\hat{\sigma}^2_p)$ whereas increasing t is most effective when sampling from skewed distributions.

Summary and Conclusions

The purpose of this investigation was to provide additional evidence related to two of the guidelines formulated by Shoemaker (1973) for

performing multiple matrix sampling. In addition, the effect on the standard error terms of rerandomizing the item universe rather than not rerandomizing after all items in the universe have been used was studied. Generalizations from the results of this study must be made cautiously. Only two item universes were studied. Furthermore, the number of replications used to estimate the accuracy of the estimated values was extremely small for studies of this type. Nonetheless, the results from this study provide limited support for the following three conclusions:

1. If the item universe is rerandomized after single exhaustion has occurred, less accurate estimates of μ and σ^2 will be obtained than if the item universe is not rerandomized.
2. If $tk>K$, more accurate estimates of μ and σ^2 are not necessarily obtained when tk is an integer multiple of K than when tk is not an integer multiple of K.
3. If $tk>K$ and if estimates of both μ and σ^2 are needed, there are few general guidelines for selecting values of t and k which will result in the most accurate estimates of μ and σ^2 from the same sampling plan.

References

Barcikowski, R. S. A monte carlo study of item sampling (versus traditional sampling) for norm construction. Journal of Educational Measurement, 1972, 9, 209-214.

Barcikowski, R. S. The effects of item discrimination on the standard errors of estimate associated with item-examinee sampling procedures. Educational and Psychological Measurement, 1974, 34, 231-237.

Knapp, T. R. Item Sampling: Unpublished manuscript, University of Rochester, 1972.

Lord, F. M. and Novick, M. R. Statistical Theories of Mental Test Scores. Reading: Addison-Wesley, 1968.

Moy, M. L. Y. and Barcikowski, R. S. Item sampling: optimal number of people and items. The Journal of Experimental Education, 1974, 42, 46-52.

Shoemaker, D. M. Allocation of items and examinees in estimating a norm distribution by item-sampling. Journal of Educational Measurement, 1970, 7, 123-128 (a).

Shoemaker. D. M. Item-examinee sampling procedures and associated standard errors in estimating test parameters. Journal of Educational Measurement, 1970, 7, 255-262 (b).

Shoemaker, D. M. Further results on the standard errors of estimate associated with item-examinee sampling procedures. Journal of Educational Measurement, 1971, 8, 215-220.

Shoemaker, D. M. Standard errors of estimate in item-examinee sampling as a function of test reliability, variation in item difficulty indices and degree of skewness in the normative distribution. Educational and Psychological Measurement, 1972, 32, 705-714.

Shoemaker, D. M. Principles and Procedures of Multiple Matrix Sampling. Cambridge: Ballinger Publishing Company, 1973.

TABLE 1

Sampling Plans Implemented

for the Mathematics Test

Plan	t	k	n	0	RANDOM	NREPS
1	3	12	20	720	NA*	10
2	4	9	20	720	NA	10
3	6	6	20	720	NA	10
4	9	4	20	720	NA	10
5	12	3	20	720	NA	10
6	4	12	20	960	NO	10
7	4	12	20	960	YES	10
8	12	4	20	960	NO	10
9	12	4	20	960	YES	10
10	5	12	20	1200	NO	10
11	5	12	20	1200	YES	10
12	12	5	20	1200	NO	10
13	12	5	20	1200	YES	10
14	6	12	20	1440	NO	10
15	6	12	20	1440	YES	10
16	12	6	20	1440	NO	10
17	12	6	20	1440	YES	10

* Not Applicable

TABLE 2

Sampling Plans Implemented

for the Vocabulary Test

Plan	t	k	n	O	RANDOM	NREPS
1	4	10	20	800	NA*	5
2	5	8	20	800	NA	5
3	8	5	20	800	NA	5
4	10	4	20	800	NA	5
5	4	12	20	960	NO	5
6	4	12	20	960	YES	5
7	12	4	20	960	NO	5
8	12	4	20	960	YES	5
9	10	7	20	1400	NO	5
10	10	7	20	1400	YES	5
11	4	20	20	1600	NO	5
12	4	20	20	1600	YES	5
13	20	4	20	1600	NO	5
14	20	4	20	1600	YES	5
15	8	10	20	1600	NO	5
16	8	10	20	1600	YES	5
17	10	8	20	1600	NO	5
18	10	8	20	1600	YES	5

* Not Applicable

TABLE 3

Results of Simple Random Assignment of Items to Subtests - Mathematics

PLAN	t/k/n	0	$\bar{\hat{\mu}}_p$	$SE(\hat{\mu}_p)$	Ψ_1	$\bar{\hat{\sigma}}^2_p$	$SE(\hat{\sigma}^2_p)$	Ψ_2	RAND	NREPS
1	3/12/20	720	11.745	.691	.576	28.486	10.306	9.128	NA*	10
2	4/09/20	720	11.595	.498	.450	28.150	5.147	4.073	NA	10
3	6/06/20	720	11.495	.542	.465	30.735	11.858	9.717	NA	10
4	9/04/20	720	11.320	.586	.489	31.066	5.512	4.239	NA	10
5	12/03/20	720	11.620	.615	.399	29.257	9.342	7.346	NA	10
6	4/12/20	960	11.535	.613	.487	25.956	4.267	5.221	NO	10
7	4/12/20	960	11.400	.830	.718	23.073	7.486	8.551	YES	10
8	12/04/20	960	11.674	.506	.394	32.040	9.721	7.857	NO	10
9	12/04/20	960	11.486	.500	.453	27.692	4.956	4.319	YES	10
10	5/12/20	1,200	11.478	.359	.329	27.004	4.280	4.907	NO	10
11	5/12/20	1,200	11.655	.666	.541	30.244	6.485	5.748	YES	10
12	12/05/20	1,200	11.901	.527	.496	27.063	3.867	4.134	NO	10
13	12/05/20	1,200	11.604	.628	.537	22.852	7.370	8.762	YES	10

TABLE 3 Cont.

14	6/12/20	1,440	11.900	.654	.615	27.638	4.643	4.363	NO	10
15	6/12/20	1,440	11.457	.778	.717	27.369	6.434	6.343	YES	10
16	12/06/20	1,440	11.747	.434	.378	27.445	5.666	4.658	NO	10
17	12/06/20	1,440	11.520	.394	.330	25.304	6.900	7.012	YES	10

* NA = Not Applicable

Data Base Values: $\mu = 11.623$, $\sigma^2 = 30.385$

TABLE 4

Results of Simple Random Assignment of Items to Subtests - Vocabulary

PLAN	t/k/n	0	$\bar{\hat{\mu}}_p$	$SE(\hat{\mu}_p)$	Ψ_1	$\bar{\hat{\sigma}}^2_p$	$SE(\hat{\sigma}^2_p)$	Ψ_2	RAND	NREPS
1	4/10/20	800	22.080	1.255	1.236	88.964	11.717	9.435	NA*	5
2	5/08/20	800	22.980	.596	.524	82.923	6.245	6.860	NA	5
3	8/05/20	800	22.900	.540	.524	93.776	8.791	7.318	NA	5
4	10/04/20	800	22.280	.682	.716	102.818	6.802	14.490	NA	5
5	4/12/20	960	23.108	.632	.599	93.628	5.699	6.278	NO	5
6	4/12/20	960	22.808	1.426	1.355	99.453	10.471	11.996	YES	5
7	12/04/20	960	23.117	1.031	.864	82.496	6.109	7.504	NO	5
8	12/04/20	960	22.833	.340	.297	104.341	14.658	16.013	YES	5
9	10/07/20	1,400	23.120	.566	.561	91.802	11.164	8.274	NO	5
10	10/07/20	1,400	23.126	.821	.852	98.802	10.752	13.163	YES	5
11	4/20/20	1,600	21.990	.597	.692	104.234	8.806	15.906	NO	5
12	4/20/20	1,600	22.745	1.212	1.109	96.293	14.319	13.289	YES	5
13	20/04/20	1,600	22.480	.054	.202	103.344	8.814	15.017	NO	5

TABLE 4 Cont.

14	20/04/20	1,600	22.685	.733	.586	86.289	7.333	6.541	YES	5
15	8/10/20	1,600	23.285	.858	.686	91.239	8.821	7.132	NO	5
16	8/10/20	1,600	22.705	.909	.839	98.749	7.015	10.421	YES	5
17	10/08/20	1,600	22.665	.668	.641	91.910	4.901	4.781	NO	5
18	10/08/20	1,600	23.090	.351	.431	95.412	10.560	10.108	YES	5

* NA = Not Applicable

Data Base Values: $\mu = 22.682$, $\sigma^2 = 88.328$

TABLE 5

Number of Times Rerandomization or Nonrerandomization of Items After Single Exhaustion Resulted in Lower Criterion Values

Test		Criterion			
		$SE(\hat{\mu}_p)$	Ψ_1	$SE(\hat{\sigma}^2_p)$	Ψ_2
Mathematics	RERAND	2	1	1	1
	NONRERAND	4	5	5	5
Vocabulary	RERAND	2	2	3	2
	NONRERAND	5	5	4	5

TABLE 6

Results for the Mathematics and Vocabulary Tests

by Sampling Plans with t< k and t>k

t < k

MATH			VOCABULARY		
Plan	$SE(\hat{\mu}_p)$	Ψ_1	Plan	$SE(\hat{\mu}_p)$	Ψ_1
1 (tk=K)	.691	.576	1 (tk=K)	1.255	1.236
2 (tk=K)	.498	.450	2 (tk=K)	.596	.524
6 (tk≠aK)	.613	.487	5 (tk≠aK)	.632	.599
10 (tk≠aK)	.359	.329			
			11 (tk=2K)	.597	.692
14 (tk=2K)	.654	.615	15 (tk=2K)	.773	.586
Plan	$SE(\hat{\sigma}^2_p)$	Ψ_1	Plan	$SE(\hat{\sigma}^2_p)$	Ψ_1
1 (tk=K)	10.306	9.128	1 (tk=K)	11.717	9.435
2 (tk=K)	5.147	4.073	2 (tk=K)	6.245	6.860
6 (tk≠aK)	4.267	5.221	5 (tk≠aK)	5.699	6.278
10 (tk≠ak)	4.280	4.907			
			11 (tk=2K)	8.806	15.906
14 (tk=2K)	4.643	4.363	15 (tk=2K)	8.821	7.132

TABLE 6 Cont.

t>k

MATH

Plan	$SE(\hat{\mu}_p)$	V_1
4 (tk=K)	.586	.489
5 (tk=K)	.615	.399
8 (tk≠aK)	.506	.394
12 (tk≠aK)	.527	.496
16 (tk=2K)	.434	.378

Plan	$SE(\hat{\sigma}^2_p)$	V_2
4 (tk=K)	5.512	4.239
5 (tk=K)	9.342	7.346
8 (tk≠aK)	9.721	7.857
12 (tk≠aK)	3.867	4.134
16 (tk=2K)	5.666	4.658

VOCABULARY

Plan	$SE(\hat{\mu}_p)$	V_1
3 (tk=K)	.540	.524
4 (tk=K)	.682	.716
7 (tk≠aK)	.340	.297
9 (tk≠aK)	.566	.561
13 (tk=2K)	.054	.202
17 (tk=2K)	.668	.641

Plan	$SE(\hat{\sigma}^2_p)$	V_2
3 (tk=K)	8.791	7.318
4 (tk=K)	6.802	14.490
7 (tk≠aK)	6.109	7.504
9 (tk≠aK)	11.164	8.274
13 (tk=2K)	8.814	15.017
17 (tk=2K)	4.901	4.781